10-2017

Efficient and robust emergence of norms through heuristic collective learning

Jianye HAO

Jun SUN

Guangyong CHEN

Zan WANG

Chao YU

See next page for additional authors

Author

Jianye HAO, Jun SUN, Guangyong CHEN, Zan WANG, Chao YU, and Zhong MING

Efficient and Robust Emergence of Norms through Heuristic Collective Learning

JIANYE HAO, School of Computer Software, Tianjin University
JUN SUN, Singapore University of Technology and Design
GUANGYONG CHEN, The Chinese University of Hong Kong
ZAN WANG, School of Computer Software, Tianjin University
CHAO YU, Dalian University of Technology
ZHONG MING, Shenzhen University

In multiagent systems, social norms serves as an important technique in regulating agents' behaviors to ensure effective coordination among agents without a centralized controlling mechanism. In such a distributed environment, it is important to investigate how a desirable social norm can be synthesized in a bottom-up manner among agents through repeated local interactions and learning techniques. In this article, we propose two novel learning strategies under the collective learning framework, *collective learning EV-l* and *collective learning EV-g*, to efficiently facilitate the emergence of social norms. Extensive simulations results show that both learning strategies can support the emergence of desirable social norms more efficiently and be applicable in a wider range of multiagent interaction scenarios compared with previous work. The influence of different topologies is investigated, which shows that the performance of all strategies is robust across different network topologies. The influences of a number of key factors (neighborhood size, actions space, population size, fixed agents and isolated subpopulations) on norm emergence performance are investigated as well.

Categories and Subject Descriptors: I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems*

General Terms: Algorithms, Experimentation

Additional Key Words and Phrases: Norm emergence, multiagent collective learning

ACM Reference format:
Jianye Hao, Jun Sun, Guangyong Chen, Zan Wang, Chao Yu, and Zhong Ming. 2017. Efficient and Robust Emergence of Norms through Heuristic Collective Learning. *ACM Trans. Auton. Adapt. Syst.* 12, 4, Article 23 (October 2017), 20 pages.
https://doi.org/10.1145/3127498

This work is partially supported by Tianjin Research Program of Application Foundation and Advanced Technology (No.: 16JCQNJC00100), National Natural Science Foundation of China (No.: 71502125, 61672358, 61502072), Fundamental Research Funds for the Central Universities of China (No.: DUT16RC(4)17).
Authors' addresses: J. Hao and Z. Wang (Corresponding authors), School of Computer Software, Tianjin University, Tianjin, China; emails: {jianye.hao, wangzan}@tju.edu.cn; J. Sun, Pillar of ISTD, Singapore University of Technology and Design, Singapore; email: sunjun@sutd.edu.sg; G. Chen, Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China; email: gychen@cuhk.edu.hk; C. Yu, School of Computer Science and Technology, Dalian University of Technology; email: cy496@dlut.edu.cn; Z. Ming, School of Computer Science and Software Engineering, Shenzhen University; email: mingz@szu.edu.cn.

https://doi.org/10.1145/3127498

1 INTRODUCTION

In multiagent systems (MASs), social norms (conventions) play an important role in regulating agents' behaviors to ensure effective coordination among agents. Social norms have been successfully applied in a wide variety of practical distributed systems such as electronic institutions (Criado et al. 2011), norm-supported computational societies (Alexander et al. 2009), and normative ad-hoc networks (Alexander et al. 2005).

There usually exist two major approaches of obtaining a social norm: the top-down approach (Morales et al. 2013; Agotnes and Wooldridge 2010) and the bottom-up approach (Sen and Airiau 2007; Yu et al. 2013, 2016; Yang et al. 2016). The top-down approach investigates how to efficiently synthesize an effective social norm for all agents beforehand, while the bottom-up approach focuses on investigating how an effective social norm can emerge through repeated local interactions among agents. In distributed multiagent environments, since there does not exist a centralized controller and the desirable norm of the system may vary when the environment changes, which is unpredictable beforehand, it is not feasible to precompute any norm for agents to employ before their interaction starts. Thus the bottom-up approach seems to be more suitable in this kind of distributed and dynamic environment. Investigating what kind of mechanism can facilitate agents towards a consistent and desirable social norm through local interaction is important to ensure the effective coordination among agents and overall high performance of the system. One significant observation is that a norm (convention) can be defined as an equilibrium that everyone is expected to follow during interactions, whereas multiple equilibria may coexist, and norms can be evolved through local learning (Young 1996; Sen and Airiau 2007). Following this observation, a natural research question to ask is how a desirable social norm can emerge in a bottom-up manner within a given (local) interaction context through the way of local learning?

Until now, much effort has been devoted by researchers from the normative MASs area to investigating the emergence of norms in agent societies through different manners of learning. Sen and Airiau (2007) first proposed applying a number of existing multiagent learning algorithms to the *social learning framework* to investigate the emergence of social norms through learning in a population of agents. However, their social learning framework is based on the assumption of random interaction and does not consider the underlying topology, which may not be able to accurately reflect the actual interaction patterns in the real world. Later, a great deal of work (Sen and Sen 2010; Villatoro et al. 2009, 2011; Yu et al. 2013) extended this social learning framework by taking into consideration complex networks (e.g., small-world and scale-free network) to model the underlying topology of the agent society, and a number of learning strategies and mechanisms have been proposed to facilitate the convergence to social norms through local interaction.

One commonly adopted abstraction of a norm in previous work is that it corresponds to a Nash equilibrium where all agents choose identical actions, which usually can be modeled as a coordination game (see Figure 1(a)). This kind of abstraction covers a number of practical scenarios, such as distributed robots coordinating on which target object to work on together and wireless nodes coordinating on which channel to reserve for control message (Mihaylov et al. 2014). However, it is often the case that the norms in practical scenarios correspond to other forms of Nash equilibria requiring agents to choose different actions, which cannot be (efficiently) handled in previous work (Yu et al. 2013). One representative example is considering that two drivers arrive at a road intersection simultaneously from two neighboring roads (Sen and Airiau 2007). To avoid collisions, one feasible norm is that each driver always yields to the driver on his/her left-hand side. This kind of scenario can be modeled as a two-player anti-coordination game (see Figure 1(b)), in which the norms correspond to the Nash equilibria (i.e., (a, b) and (b, a)) with complementary actions for each player. Furthermore, more complicated scenarios involve multiple Nash equilibria

1's payoff, 2's payoff		Player 2's action	
		a	b
Player 1's action	a	1,1	-1,-1
	b	-1,-1	1,1

(a)

1's payoff, 2's payoff		Player 2's action	
		a	b
Player 1's action	a	-1,-1	1,1
	b	1,1	-1,-1

(b)

1's payoff, 2's payoff		Player 2's action		
		a	b	c
Player 1's action	a	10,10	0,0	-30,-30
	b	0,0	7,7	0,0
	c	-30,-30	0,0	10,10

(c)

1's payoff, 2's payoff		Player 2's action		
		a	b	c
Player 1's action	a	12/8, 12/8	5/-5,5/-5	-40/-20, -40/-20
	b	5/-5,5/-5	14/0, 14/0	5/-5,5/-5
	c	-40/-20, -40/-20	5/-5,5/-5	12/8, 12/8

(d)

Fig. 1. Payoff matrices for (a) coordination game (CG), (b) anti-coordination game (ACG), (c) coordination game with high penalty (CGHP), and (d) fully stochastic coordination game with high penalty (FSCGHP).

while only some of them correspond to norms. In this kind of scenario, it is very likely for agents to converge to the non-norm equilibrium to avoid the high mis-coordination cost, and this issue has not been addressed by previous work (Sen and Airiau 2007; Yu et al. 2013) yet. One specific example is shown in Figure 1(c), in which there exist two equilibria (a, a) and (c, c) representing the norms and a non-norm equilibrium (b, b). Due to the high penalty cost for miscoordination (i.e., reaching (a, c) or (c, a)), it is very likely for agents to converge to the non-norm equilibrium (b, b). Last but not least, it would be more challenging for the norm(s) to evolve if the interaction environment becomes stochastic. One representative example is shown in Figure 1(d), which is a stochastic version of the CGHP game in Figure 1(c). This game shares the same norm with the deterministic version in Figure 1(c), except that the payoff for each outcome is non-deterministic due to the stochasticity of the interaction environment.

To tackle the above challenges, we propose two novel learning strategies: *collective learning EV-l* and *collective-learning EV-g*, under the *networked collective learning framework*, which is applicable to a wide variety of scenarios for norm emergence. Under the *networked collective learning framework*, there are a population of agents where each agent is allowed to interact with its neighbors determined by the underlying network topology. The interaction between each pair of agents is modeled as a two-player m-action strategic game. During an interaction, each agent is assigned randomly to be the row or column player. During each round, each agent first learns and estimates the best action to play with each neighbor separately and then synthesizes an overall best-response action to interact with all neighbors. In *collective learning EV-l*, the best-response action is synthesized based on local exploration while it is based on global exploration in *collective learning EV-g*. Besides, to overcome the possible side-effects of high mis-coordination cost and uncertainty of the environment, we propose that each agent's learning strategy should incorporate both the optimistic assumption and the relative frequency information of each action based on its experience with its neighbors. At the end of each round, each agent updates its learning strategy against each

neighbor based on its past experience accordingly. We extensively evaluate the learning performance of both *collective learning EV-l* and *collective learning EV-g* and show that both strategies enable agents to reach consistent norms more efficiently in a wider range of games than previous approaches. We also empirically find that the performance of *collective learning EV-l* and *EV-g* is robust across different topologies. The influence of different parameters and factors (neighborhood size, population size, action space, fixed agents and isolated subpopulation) on the learning performance of *collective learning EV-l* and *EV-g* are also investigated.

The remainder of the article is organized as follows. In Section 2, the networked collective learning framework and the collective learning strategy are described. In Section 3, we present the learning performance of *collective learning EV-l* and *collective learning EV-g* under the networked collective learning framework in different types of interaction scenarios by comparing it with previous work, and also investigate the influence of different parameters. In Section 4, we give a brief overview of previous work on norm emergence in MASs. Last, Section 5 concludes with pointing out future directions.

2 NETWORKED COLLECTIVE LEARNING FRAMEWORK

Under the networked collective learning framework, there are a population of agents in which each agent's neighbors are determined by the underlying network topology. Three major topologies are considered: ring network, small-world network (Réka and Barabási 2002), and scale-free network (Barabási et al. 2000). Each agent i learns its policy (i.e., which norm to adopt) through repeated pairwise interactions with all its neighbors each round. The interaction between each pair of agents is modeled as a two-player, m-action stage game. These stage games typically have pure strategy Nash equilibria. In each round, each agent is paired with a randomly selected agent from its neighborhood to interact. Following the setting in previous work (Sen and Airiau 2007), one agent is randomly chosen as the row player and the other agent as the column player during each interaction. The agents are assumed to know their roles (states), that is, either as row player or column player, before the start of each interaction.

The overall networked collective learning framework is presented in Algorithm 1. At the beginning of each round t, each agent first determines the set $\mathcal{S}_{i,r}^t$ and $\mathcal{S}_{i,c}^t$ of its current best-response policy against each of its neighbors as either the row and column player, respectively (line 3). Formally, we have

$$\mathcal{S}_{i,r}^t = \left\{P_{i,1}^{t,r}, P_{i,2}^{t,r}, \ldots, P_{i,N(i)}^{t,r}\right\} \tag{1}$$

and

$$\mathcal{S}_{i,c}^t = \left\{P_{i,1}^{t,c}, P_{i,2}^{t,c}, \ldots, P_{i,N(i)}^{t,c}\right\}, \tag{2}$$

where $N(i)$ is agent i's neighborhood size and $P_{i,k}^{t,r}$ and $P_{i,k}^{t,c}$ are agent i's current best response policy towards its neighboring agent k when it is the row and column player, respectively.

Following that, the sets $\mathcal{S}_{i,r}^t$ and $\mathcal{S}_{i,c}^t$ of best response actions are synthesized into a single best response policy $P_{i,*}^{t,r}$ and $P_{i,*}^{t,c}$, respectively (line 4), which will be used as the overall strategy to interact with all of its neighbors in the current round of interaction t. This models people's collective decision-making process in which people make collective decisions based on multiple feedbacks. How the agents determine the best responses against their neighbors and synthesize the overall best-response actions will be described in detail in Section 2.2. In each round, each agent i has the opportunity to interact with each of its neighbors once. The interaction between each agent and each of its neighbors is modeled as a two-player stage game, in which their roles are assigned randomly (line 6–11). During each interaction, each agent uses its best response policy ($P_{i,*}^{t,r}$ or $P_{i,*}^{t,c}$) to play the game with its partners and receives a reward accordingly. It is assumed

ALGORITHM 1: Overall Description of the Networked Collective Learning Framework

1: **for** a number of rounds **do**
2: **for** each agent i **do**
3: Determine its best-response policy set $\mathcal{S}_{i,r}^t$ and $\mathcal{S}_{i,c}^t$ respectively.
4: Synthesize its single best-response policy as the row and column player $P_{i,*}^{t,r}$ and $P_{i,*}^{t,c}$ based on $\mathcal{S}_{i,r}^t$ and $\mathcal{S}_{i,c}^t$.
5: **end for**
6: **for** each agent i **do**
7: **for** each neighboring agent j **do**
8: Play a stage game under role s (randomly assigned)
9: Update its learning strategy against each neighbor j using $\langle P_{i,*}^{t,s}, R_{i,j}^t \rangle$.
10: **end for**
11: **end for**
12: **end for**

that each agent can only perceive the action and payoff of its own during each interaction. At the end of each interaction, each agent updates its learning strategy towards each neighbor based on its current round experience (line 9).

2.1 Network Topology

We focus on three representative network topologies (ring network, small-world network, and scale-free network) in this article and briefly describe the characteristic of each topology in this section (Wang and Chen 2003).

— *Ring network:* As a widely studied and regular network, in a typical ring network, each node connects to k nearest-neighbor nodes on each side. Each node shares the same connectivity degree in a ring network. The diameter and average path length of the ring network is increased with the size N of the network and goes to infinity as $N \to \infty$. The clustering coefficient of the ring network is increased with the connectivity degree k, and the network becomes fully connected eventually. We denote a ring network as RW_N^k, where N is the size of the network and k is the connectivity degree of the network.

— *Small-world network:* In a typical small-world network, its connectivity (degree) distribution peaks at an average value and decays exponentially on both sides, that is, the connection degree of most nodes is the same. This kind of networks is also featured by high clustering coefficients and short average path lengths. We denote a small-world network as $SW_N^{k,\rho}$, where N is the size of the network, k is its average connectivity degree, and ρ is the re-wiring probability indicating the degree of the network randomness. The small-world network reflects the "what a small world" phenomenon reflected in many practical networks such as collaboration networks (e.g., the co-authorship of research articles) and the social influence networks (Réka and Barabási 2002).

— *Scale-free network:* Different from small-world networks, the connectivity distribution of a scale-free network follows the power-law distribution, that is, for each node, the probability of being connected to k adjacent nodes is proportional to $k^{-\gamma}$ (γ is a constant). Intuitively this indicates most of the nodes have very few connections while only a few nodes have very large connections. This kind of "scale-free" feature has been observed in many real-world networks such as the connection network of web pages (Barabási et al. 2000) and citation

network of research articles (Redner 1998). A scale-free network can usually be denoted as SF_N^γ, where N represents the network size.

2.2 Collective Learning Strategy

In general, to achieve coordination on efficient norms, an agent's behavior varies depending on its current role (its perceived state). For example, considering the problem of "rule of the road" in which two drivers arrive at an intersection simultaneously from two neighboring roads and decide which one yields to avoid collision (Sen and Airiau 2007). In this case, one desirable norm is that the agent on the right side always yields to agents coming from its left side. This requires any pair of encountering agents to choose different actions, and thus a suitable coordination policy must be a set of policies specifying an action for each state to avoid any possible miscoordinations. For example, one intuitive policy would be "always yielding to the car on the left-hand side", which can be represented as "Stop when the neighboring car is on your left-hand side and move when the neighboring car is on you right-hand side."

Formally, we propose that each agent i holds a Q-value $Q_{i,j}(s,a)$ for each action a under each state $s \in \{Row, Column\}$ against each of its neighbors j, which keeps a record of action a's past performance against neighbor j and serves as the basis for making decisions. Given the learning information at the end of each round t, each agent i updates its Q-values for each neighbor j following Equation (3),

$$Q_{i,j}^{t+1}(s,a) = Q_{i,j}^{t}(s,a) + \alpha_{i,j}^{t}(s) \times \left[R_{i,j}^{t}(s,a) - Q_{i,j}^{t}(s,a)\right], \tag{3}$$

where $R_{i,j}^t(s,a)$ is agent i's immediate reward in the current round when interacting with its neighbor j by choosing action a and $\alpha_{i,j}^t(s)$ is its current learning rate in state s for its neighbor j.

There are three major factors that may influence an agent's success rate of converging to norms. One is its neighbor agents' behaviors, which is changing dynamically and usually leads to low payoff due to the mis-coordination among agents. Thus the agents are likely to be motivated to suboptimal norms to avoid high mis-coordination cost through learning. One specific strategic game modeling this kind of situation is shown in Figure 1(c), in which the agents are very liable to converge to the suboptimal outcome (b,b) due to the high mis-coordination cost of -30. To deal with the influence of the neighbor agents' behaviors, we update the Q-value in an optimistic manner by ignoring the penalty due to mis-coordination. Another factor is the stochastic property of the environment, which leads to the problem of how the agents should distinguish between the exploration of the opponents and the stochasticity of the environment. We can handle the stochasticity of the environment by taking the relative frequencies of different rewards of an action into consideration during the Q-value update. The last factor is the network-based learning environment in which each agent interacts with multiple interaction partners each round and may learn towards different directions (different optimal actions may be learned for different partners), thus impeding the emergence of a consistent norm in the system. We can handle this issue by employing synthesis strategies for each agent to come up with a single action to interact with all interaction partners.

To this end, we employ the Frequency Maximum Q-value heuristic (FMQ) (Kapetanakis and Kudenko 2002) as the updating heuristic to compute the estimated values of the actions. The FMQ heuristic was proposed to overcome the mis-coordination problem in fixed-agent repeated interaction framework. This heuristic is not only based on the optimistic assumption but also takes into consideration the relative frequency of the maximum reward being received for each action. In the networked collective learning framework, since each agent interacts with all its neighbors each round, we propose applying the FMQ heuristic on the combination of its past history and its current round experience with all neighbors.

Specifically, for each agent i, let us denote its learning experience in the current round t as $P_i^t = \{\langle a_{i,1}^t, R_{i,1}^t\rangle, \ldots, \langle a_{i,N(i)}^t, R_{i,N(i)}^t\rangle\}$, where $N(i)$ is its neighborhood size. Also let us classify P_i^t into two distinct subsets $P_{i,r}^t$ and $P_{i,c}^t$ based on the role of the agent i (namely *Row* (r) or *Column* (c) player) during each interaction. Based on the FMQ heuristic, each agent i assesses the relative performance $EV_{i,j}(s,a)$ of each action a against the neighboring agent j under the current state s as follows:

$$EV_{i,j}^{t+1}(s,a) = Q_{i,j}^{t+1}(s,a) + c \times f^{t+1}(s,a) \times R_{max}^t(s,a), \tag{4}$$

where

- $R_{max}^t(s,a) = \max\{R \mid \langle a, R\rangle \in P_{i,s}^t\}$, $s \in \{r,c\}$, which is the highest payoff in the set $P_{i,s}^t$,
- $f^{t+1}(s,a)$ is the frequency of receiving the reward of $R_{max}^t(s,a)$ until now by choosing action a,
- c is the weighting factor determining the relative importance of $R_{max}^t(s,a)$.

The value of $f^{t+1}(s,a)$ is obtained based on the combination of the past history and the current-round experience with all neighbors and is computed as follows:

$$f^{t+1}(s,a) = f^t(s,a) + \frac{1}{t+1}[f_s^t(s,a) - f^t(s,a)], \tag{5}$$

where $f_s^t(s,a)$ is the average frequency of receiving the reward of $R_{max}^t(s,a)$ by choosing action a based on the current round experience $P_{i,s}^t$ only.

Based on its corresponding set of EV-values, each agent chooses its best response action against each neighbor under each state using the ϵ-greedy mechanism. Specifically, each agent chooses its action with the highest EV-value with probability $1-\epsilon$ (randomly selection in case of a tie) to exploit the action with best estimated performance currently and makes random choices with the rest of probability ϵ to ensure that those actions with potentially better performance have the opportunity to be explored.

Finally, each agent synthesizes a single best-response action for both roles based on the sets $\mathcal{S}_r^t$ and $\mathcal{S}_c^t$ of best-response actions against each neighbor under the row and column role. This synthesis process imitates people's collective decision-making process in which people usually consult with multiple alternative opinions before making the final decision (Polikar 2006). Given $\mathcal{S}_r^t = \{a_i^t(1), a_i^t(2), \ldots, a_i^t(N(i))\}$ and $\mathcal{S}_c^t = \{b_i^t(1), b_i^t(2), \ldots, b_i^t(N(i))\}$, respectively, each agent synthesizes the overall best-response actions under both roles based on the majority voting. The relative preference $p_i^t(s,a)$ of each action a in state $s \in \{r,c\}$ is determined by the number of times that this action is selected as the best-response action against different neighbors. Formally, it can be represented as follows:

$$p_i^t(s,a) = \begin{cases} \sum_1^{N(i)} I(a, a_i^t(i)) & \text{if } s = r \\ \sum_1^{N(i)} I(a, b_i^t(i)) & \text{if } s = c \end{cases}, \tag{6}$$

where $N(i)$ is the neighborhood size of agent i, and $I(a, a_i^t(i))$ and $I(a, b_i^t(i))$ represent the indicator function defined as follows:

$$I(a,a') = \begin{cases} 1 & \text{if } a = a' \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

Agent i's overall best-response actions in round t (namely $a_i^{t,*}$ and $b_i^{t,*}$) are determined as the actions with the highest preferences under the row (r) and column (c) states, respectively. At the early stage of learning, the agents may not have an accurate estimation of the EV-value of each action against each neighbor, and thus the synthesized actions may not be the optimal best-response

actions. It is necessary for the agents to make additionally random explorations initially to explore those actions with possibly better performance. We distinguish two general ways of making explorations: local exploration and global exploration.

Local exploration. The exploration is made before synthesizing the overall best-response action. Each agent makes explorations to determine the best-response action against each neighbor in each state based on the ϵ-greedy mechanism as previously mentioned. The synthesized best-response action is always selected as the action with the highest preference. We denote the collective learning strategy with local exploration as *collective learning EV-l*.

Global exploration. The exploration is made after the best-response action has been synthesized. The best-response action against each neighbor under each state is always determined as the action with the highest EV-value without the ϵ-greedy exploration. After that, each agent synthesizes the overall best-response action and then makes explorations following the ϵ-greedy mechanism. We denote the collective learning strategy with global exploration as *collective learning EV-g*.

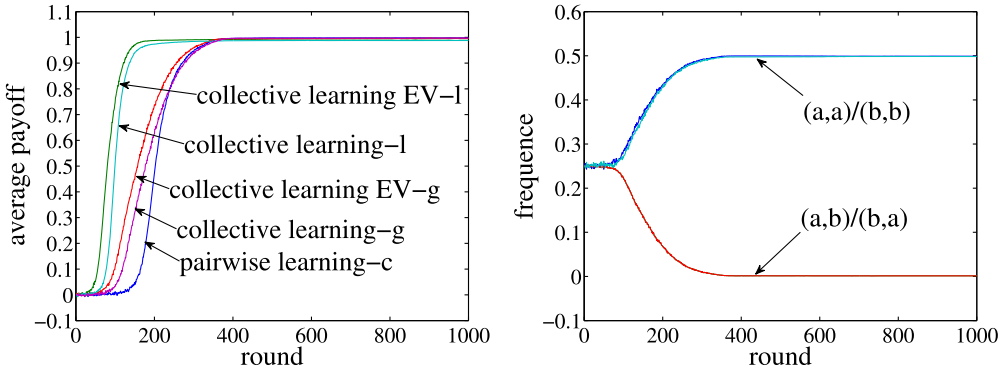
3 EXPERIMENTAL SIMULATION

3.1 Performance Evaluation

We evaluate the performance of both *collective learning EV-l* and *collective learning EV-g* in different types of interaction scenarios by comparing it with previous work: collective learning-l and g (Yu et al. 2013, 2014) and pairwise learning (Sen and Airiau 2007). Pairwise learning only allows each agent to interact with one randomly selected neighbor each round. To make the comparison fair, we modify pairwise learning to allow each agent to interact and learn with all neighbors each round, which we denote as *pairwise learning-c*. Thus all learning strategies we compare here are evaluated under the same framework.

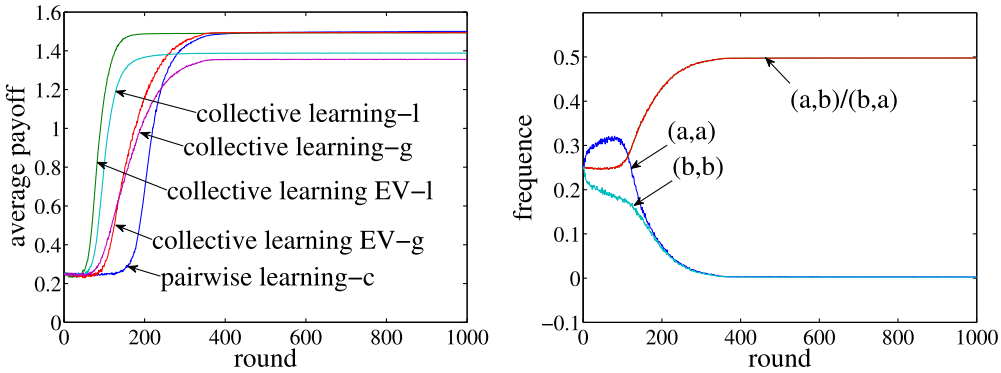
Unless mentioned otherwise, the initial learning rate α and the initial exploration rate ϵ are set to 0.8 and 0.9, respectively, which are all decreased exponentially ($\alpha/\epsilon = \alpha_{init}/\epsilon_{init} * 0.9^t$). The initial Q-values are randomly generated within the range of [0,1]. Another key parameter is the weighting factor c, which reflects the optimistic degree of an agent's updating strategy. We have extensively analyzed the influence of different values of c on the norm emergence performance and found that there would be no significant performance increase when c becomes larger than 10 across all games. Thus, in the rest of all simulation results, the weighting factor c is set to 10. The average connection degree of all networks is set to 6. All experiments are conducted within a population of 100 agents in the small-world network, and all results are averaged over 1,000 runs. The influence of different network topologies will be discussed in Section 3.2 in detail.

Coordination Game (CG). We start with the simplest testbed of the coordination game (Figure 1(a)) adopted in Yu et al. (2013), in which there exist two different norms: (a,a) and (b,b). This game represents the "rule of the road" scenario where two cars decide which side of road to drive, and two norms correspond to either driving on the left or on the right. Figure 2(a) shows the dynamics of the average payoff of agents as the function of the number of rounds for different learning strategies. We can see that all strategies enable agents to coordinate towards achieving an average payoff of 1. Figure 2(b) shows the average frequency of converging to each outcome under collective learning EV-l over 1000 runs. We can see that both norms ((a, a) and (b, b)) can be reached with equal frequency (converged in the same number of runs). This is reasonable, since the coordination game itself is symmetric, and there is no difference between these two norms. Besides, collective learning EV-l and EV-g converge faster than collective learning-l and g,



(a) CG (b) Frequency

Fig. 2. The dynamics of the average payoffs of agents in coordination game under different approaches.



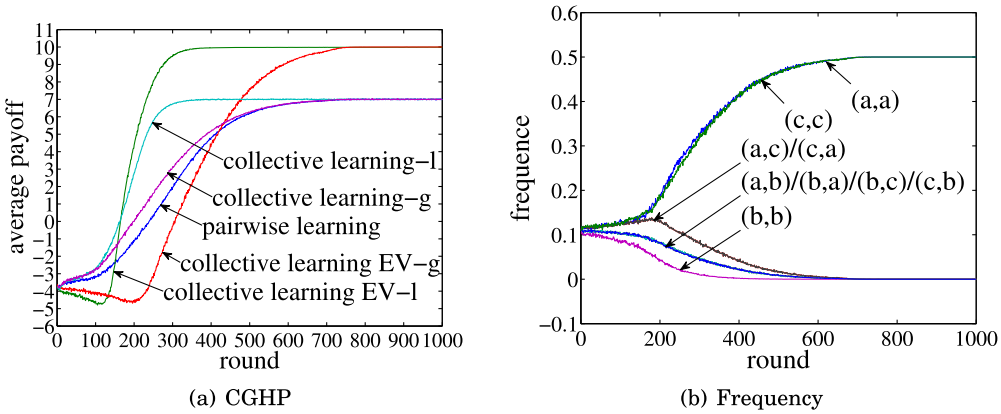
(a) ACG (b) Frequency

Fig. 3. The dynamics of the average payoffs of agents in anti-coordination games under different approaches.

respectively, and pairwise learning-c is the slowest one. Another observation is that learning strategies with local explorations converge faster than strategies with global explorations.

Anti-Coordination Game (ACG). Next we consider a scenario where individuals need to choose different actions to reach a consensus on which norm to adopt, which can be naturally modeled as an anti-coordination game in Figure 1(b). One practical example is considering two cars in a road junction choosing which car should yield to another one. Both norms correspond to yielding to either the left-hand or right-hand car. We also consider another non-symmetric variant of this anti-coordination game by changing the payoff profile under (a, b) from $(2, 1)$ to $(0, 0)$. This variant game consists of one optimal norm (b, a) and one suboptimal norm (a, b) to model the preferences on different norms.

Figure 3(a) shows the dynamics of the average payoff of agents with the number of rounds averaged over the above two anti-coordination games for different learning strategies. We can observe that the agents using collective learning EV-l/EV-g and the pairwise learning-c are able to successfully achieve the highest average payoff of 1.5, while fail for both collective learning-l and g (Yu et al. 2013). Lower average payoff (about 7–11%) is achieved for both collective learning-l and g due to high miscoordination on a consistent norm. The advantage of collective EV-l/EV-g would



(a) CGHP (b) Frequency

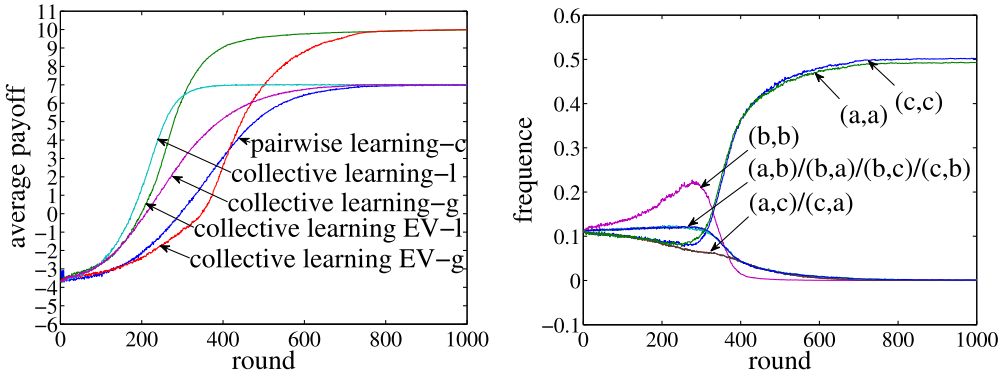
Fig. 4. The dynamics of the average payoffs of agents in CGHP under different approaches.

be more obvious when the payoff difference between successful and unsuccessful coordination are enlarged. The reason is that in collective learning-l/g, the agents cannot distinguish their roles and always adopt the same action when they are assigned as different roles. Thus mis-coordination occurs when two agents during an interaction choose the same action. Another observation is that both collective learning EV-l and EV-g converge faster than pairwise learning-c. We hypothesize that it is because agents can gain more experience under the collective learning framework and also the action synthesis mechanism enables agents to coordinate with each other more efficiently. Finally, Figure 2(b) gives the expected frequency of converging to different outcomes under collective learning EV-l averaged over 1,000 runs. We can see that all agents can eventually reach either of the norms (a, b) or (b, a) with equal frequency.

Coordination Game with High Penalty (CGHP). Next we consider one variant of the coordination game with high mis-coordination cost as shown in Figure 1(c). In this game, similarly to the coordination game, there exist two desirable norms (a, a) and (c, c). However, this game also has one suboptimal norm (b, b), to which the agents are very likely to converge due to the high mis-coordination cost of -30 when reaching (a, c) or (c, a).

Figure 4(a) shows the dynamics of the average payoff of agents using different learning strategies. From Figure 4(a), we can see that the agents can successfully reach one consistent norm (achieve the average payoff of 10) eventually using both collective learning EV-l and EV-g under the collective learning framework. In contrast, the agents fail to reach norm (actually reach the subnorm (b, b)) and only achieve the average payoff of 7 under both collective learning-l/g (Yu et al. 2013) and pairwise learning-c (Sen and Airiau 2007). The reason is that collective learning EV-l/EV-g is able to prevent agents from reaching those outcomes with high mis-coordination cost partially due to the incorporation of the optimistic assumption. For collective learning-l/g and pairwise learning-c, the agents are intimidated by the high cost of -30 when reaching either (a, c) or (c, a) due to mis-coordination at the early stage and thus converge to the non-norm outcome (b, b) eventually. Besides, Figure 2(b) illustrates the expected frequency of converging to different outcomes under collective learning EV-l in CGHP averaged over 1,000 runs. From Figure 2(b), we can see that all agents can eventually reach either of the norms (a, a) or (c, c) with equal frequency.

Fully Stochastic Coordination Game with High Penalty (FSCGHP). Finally, we consider the fully stochastic version of the CGHP shown in Figure 1(d). In FSCGHP, each outcome is associated with two possible payoffs, and the agents receive one of them with probability 0.5, which models the



(a) FSCGHP (b) Frequency

Fig. 5. The dynamics of the average payoffs of agents in FSCGHP under different approaches.

uncertainty of the interaction. This game is in essence the same with the CGHP in which there exist two desirable norms (a, a) and (b, b) but is more challenging to achieve. Figure 5(a) illustrates the dynamics of the average payoff of agents employing collective learning EV-l/EV-g and collective learning-l/g (Yu et al. 2013) and pairwise learning (Sen and Airiau 2007), respectively.

The results are similar as the CGHP: The agents employing both collective learning EV-l and EV-g are able to reach a consistent norm under the collective learning framework while fail using other learning strategies. Both collective learning-l/g and pairwise learning converge to the suboptimal outcome (b, b). Since collective learning EV-l/g takes into consideration frequency information, it is able to overcome the noises from the environment. In contrast, other approaches fail for two reasons: (1) They might mistakenly consider the outcome (b, b) as the optimal one that produces the highest payoff 14 with probability 0.5, and (2) the high mis-coordination cost of (a, c) and (c, a) makes agents stay away from reaching the outcomes (a, a) and (c, c). Last, we provide the expected frequency of converging to different outcomes under collective learning EV-l averaged over 1,000 runs in Figure 2(b). It is obvious that all agents can eventually reach either of the norms (a, a) or (c, c) with equal frequency. We notice that there is a sharp increase in the percentage of agents reaching (b, b) around 300 rounds. We hypothesize this is due to agents' inaccurate estimation of the frequency of achieving the maximum payoff by choosing action b. This inaccurate estimation lead to the consequence that the estimated EV-value of action b is temporarily higher than the rest of actions (see Equation (4)). After sufficient rounds of interactions, the frequency estimation becomes accurate, and thus the EV-values of action a (or c) becomes the highest.

Summary and Discussion. From previous results, we can see that both collective learning EV-l and EV-g are able to support norm emergence in a much wider variety of interaction scenarios than previous work (Sen and Airiau 2007; Yu et al. 2013). Besides, two important observations can be found here: (1) Agents converge to norms faster under collective learning framework than social learning framework. In other words, collective learning framework is more efficient than the social learning framework in terms of norm emergence, since under the collective learning framework each agent interacts with all neighbors each round and synthesizes all the interactions and thus learns faster. (2) Agents converge to norms faster using local exploration than global exploration under the collective learning framework. The agents are able to make more efficient explorations under the local exploration scheme and thus evolve norms faster, since the agents make independent explorations against each neighbor under local exploration scheme, while only one exploration is made under the global exploration scheme.

Table 1. Average Number of Rounds Needed Before Converging to a Consistent Norm

Average No. of Rounds	Collective learning EV-l	Collective learning EV-g	Collective learning-l	Collective learning-g	pairwise learning
Small-world	132 /138/ 277/ 443	261/ 271/ 623/ 674	133/ X/ X/ X	327/ X/ X/ X	263/ 271/ X/ X
Scale-free	131/ 141/ 287/ 486	259/ 259/ 641/ 659	144/ X/ X/ X	323/ X/ X/ X	264/ 268/ X/ X
Ring	136/ 143/ 292/ 431	265/ 272/ 638/ 674	146/ X/ X/ X	332/ X/ X/ X	268/ 268/ X/ X
Random	137/ 148/ 293/ 456	268/ 275/ 650/ 682	147/ X/ X/ X	337/ X/ X/ X	270/ 276/ X/ X

3.2 Evaluation Under Different Network Topologies

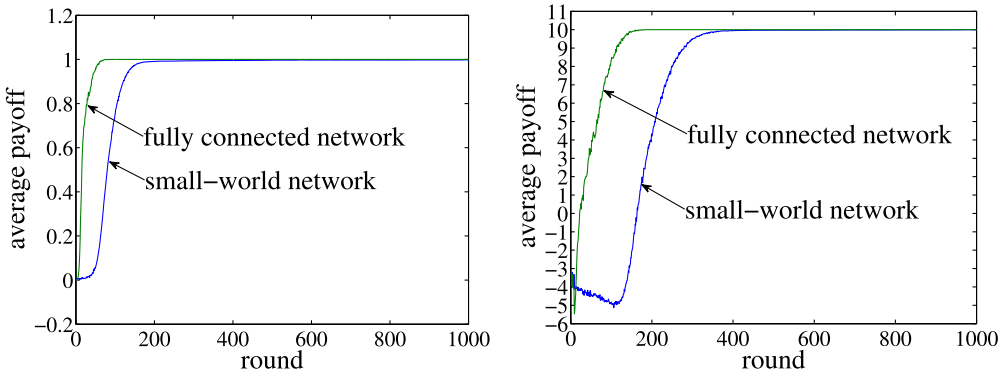
The previous section has shown the superior learning performance of both collective-learning EV-l and EV-g using the small-world network as the underlying topology. In this section, we further evaluate the performance of collective-learning EV-l and EV-g under a number of different network topologies compared with previous approaches. Three representative network topologies are considered here: *ring network*, *small-world network*, and *scale-free network*. Small-world and scale-free networks are two representative network topologies modeling real-world community structures. We also consider the case of *random network* where there is no fixed network topology and each agent simply interacts with a fixed and same number of agents randomly selected from the system each round. To make the evaluation results comparable, the number of agents that each agent interacts with in the *random network* is set to be equal to the expected connection degree (average neighbors of each agent) of the above three topologies, which are set to 6. Other parameters are set to the same as Section 3.1.

Table 1 lists the average number of rounds required before convergence for all learning strategies under different network topologies for all four types of games (each entry contains the results for the four games and "X" denotes that agents fail to converge to one consistent norm). Note that All the results are averaged over 1,000 times.

From Table 1, we can see that the results share similar patterns across all network topologies as follows:

- —Both *collective learning EV-l* and *collective learning EV-g* are always able to achieve coordination on norms for all four types of games;
- —Both *collective learning-l* and *collective learning-g* can only succeed in coordination games, while fail in the rest of games;
- —*Pairwise learning* can succeed in both coordination and AC games, while fail in the rest of games.
- —for all networks, it generally takes longer time to converge to norms (or fails to converge) when the game becomes more challenging.

From the above results, we can observe that the performance of *collective learning EV-l* and *EV-g* is robust towards different network topologies. Also *collective learning EV-l* can enable agents to converge to norms in all different types of games and more efficiently than all other approaches. Another observation is that the underlying topology itself actually has no influence on the norm emergence for all the approaches we evaluated. To evolve norms through local learning, it does



(a) Coordination Game (b) Penalty Game

Fig. 6. The dynamics of the average payoffs of agents for collective learning EV-l under the fully connected network.

not matter how the agents are actually interconnected in the system. The key factor is the way that each agent interacts with each other (determined by the learning framework) and how agents learn and make decisions through interaction (controlled by the learning strategy). The norm emergence performance is independent of the network topology and the identities of agents that they interact with (all agents are assumed to adopt the same learning strategy).

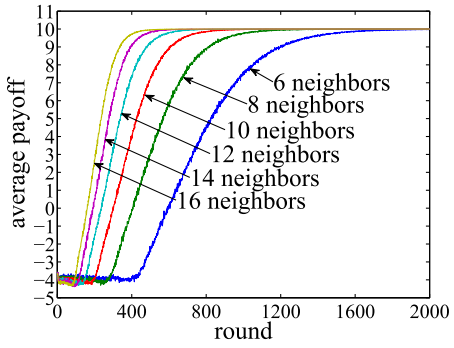
To further validate our hypothesis, we consider an extreme case—a fully connected network in which each agent is connected to all the rest of agents in the system. Each agent interacts with all other agents each round under all the previous approaches. We evaluate the norm convergence rate of all the previous approaches under the fully connected network. Simulation results show that the convergence rate is increased for all approaches we considered here. We give two representative results of collective learning EV-l under coordination game and penalty game for illustration in Figures 6(a) and (b), respectively. From both figures, we can observe that the average number of rounds needed before convergence is significantly reduced for both cases when the underlying topology is changed from a small-world network to a fully connected network.

3.3 Influence of the Size of Neighborhood, Population, and Action Space

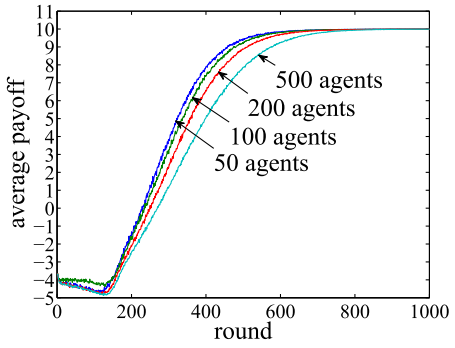
Next we evaluate the influences of other parameters on norm emergence, which are independent of the network topologies. We only present the results for *collective learning EV-l* under the small-world network averaged over coordination games. The payoffs of the coordination games are randomly generated and the results are averaged over 1,000 runs. The parameter settings follow the setting in Section 3.1 except the parameter being evaluated is changed. The results for *collective learning EV-g* are similar and omitted.

Influences of the neighborhood size. Figure 7(a) shows the dynamics of average payoffs of agents when the average neighborhood size varies. We can observe that the norm emergence rate becomes faster with the increase of the average neighborhood size. This is because the agents become more clustered with the increase of the average neighborhood size, and agents with a long distance become closer to each other. Therefore, agents need fewer interactions to reach a consistent norm when the neighborhood size is increased.

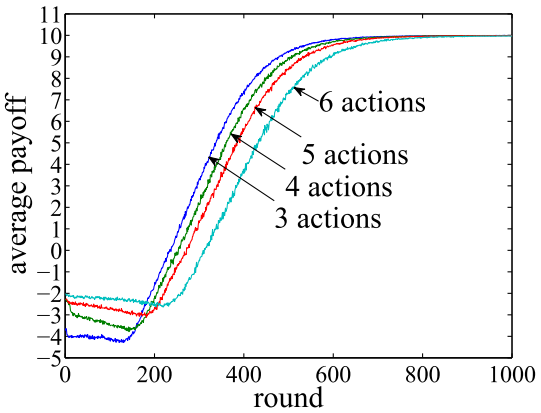
Influences of the size of the population. Figure 7(b) shows the dynamics of average payoffs of agents when the population size varies. We can see that the rate of norm convergence is delayed



(a) Influence of the neighborhood size



(b) Influence of the population size



(c) Influence of the action space

Fig. 7. The influence of different parameters on the learning performance of agents.

as the population size increases. Since agents make collective decisions based on local interactions only, the increase of the population size requires more efforts for agents to reach a consistent norm.

Influences of the number of actions. Figure 7(c) shows the dynamics of average payoffs of agents when the agents' action space varies. We can see that the increase of the action space results in the delayed convergence of norms. This is because a larger action space indicates the existence of more suboptimal actions, and the agents need more time and experience to distinguish between the optimal action and the rest of suboptimal ones before reaching a consistent norm.

3.4 Influence of Fixed Agents

From the results in Section 3.1, we can see that all norms are evolved with equal frequency over multiple runs if multiple norms coexist. This is reasonable, since all norms in the games we evaluated here are symmetric and the agents do not have any preference over different norms. In this section, we investigate the extraneous influence from outside on the direction of norm emergence. Specifically we consider injecting a small number of agents with fixed behaviors (adopting a particular norm already). We study the influence of this small amount of fixed agents on the overall norm emergence of the system. For this study, we use the CGHP game (Figure 1(c)) and consider a population of 1,000 agents using collective learning EV-l. Similar results can be observed across

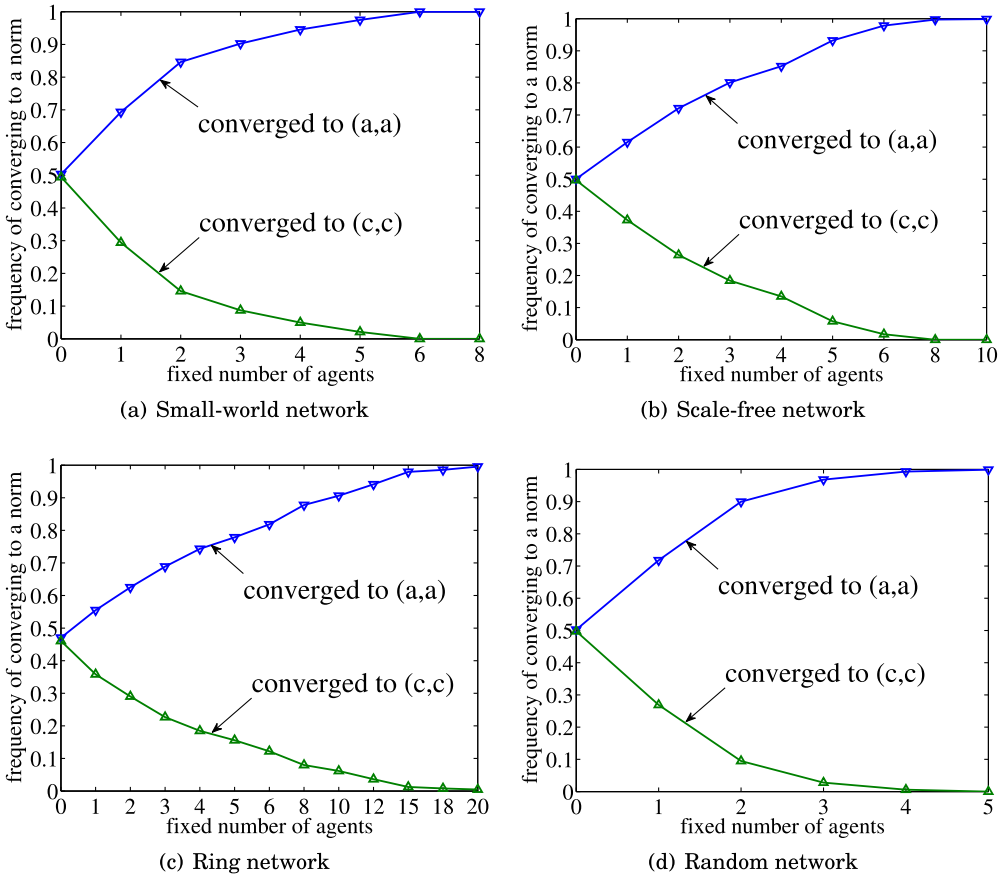


Fig. 8. Influence of fixed agents.

all the previous games we considered. The behavior of the fixed agents is adopting the norm of choosing action a under each state. The location of each fixed agent is randomly selected for all network topologies.

Figure 8 shows the frequency of the two norms are evolved when the number of fixed agents is increased gradually under the four network structures. For all four networks, initially both norms ((a, a) and (c, c)) are evolved with equal frequency. As the number of fixed agents increases, the frequency of evolving norm (a,a) is gradually increased and reaches 1 eventually. On the other hand, the number of fixed agents required for the population to converge to norm (a, a) with 100% frequency varies for different network topologies: Random > Small-world > Scale-free > Ring. Intuitively, for a random network, it requires the least number of fixed agents, since it allows interaction between any pair of agents, thus accelerating the spread speed of the fixed agents' influence to the rest of the population. In contrast, for the rest of topologies, the information transmission speed is bounded by their average path length. Besides, for small-world and scale-free networks, a fixed agent may be assigned to the node with very large number of connections (a hub), and thus it is expected to have better performance than ring networks, which can be verified from our simulation results.

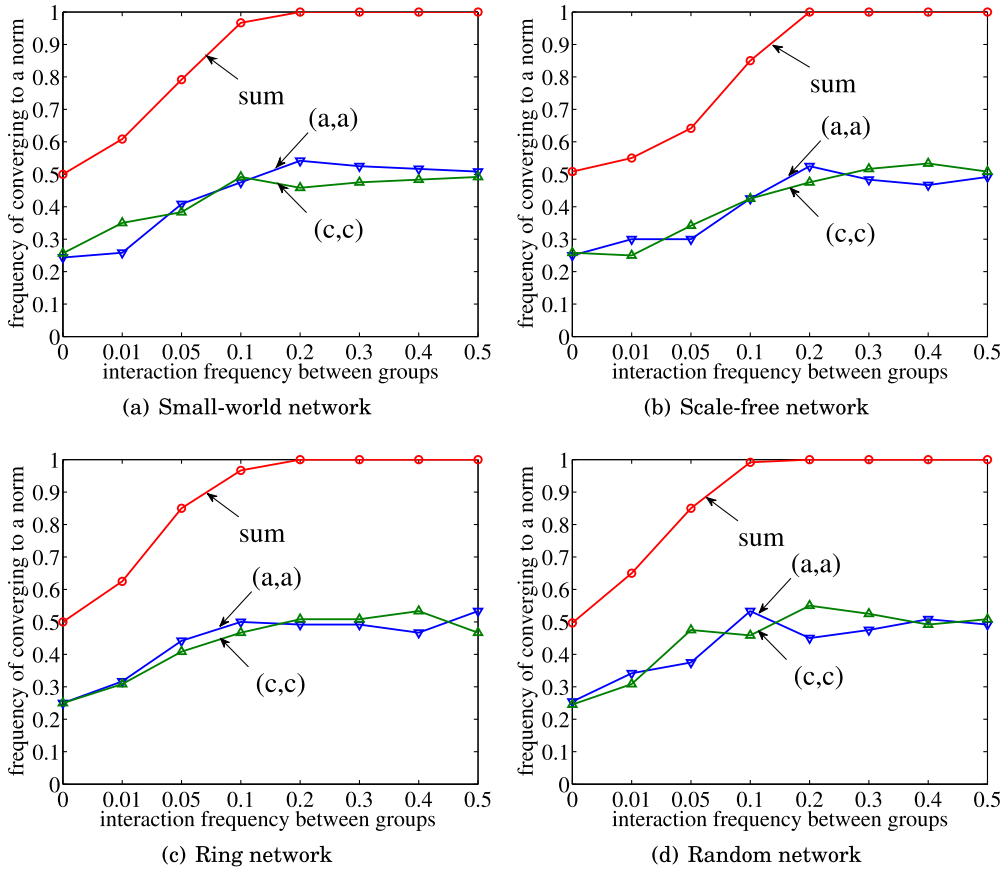


Fig. 9. Norm emergence in isolated subpopulations.

3.5 Norm Emergence in Isolated Subpopulations

In human society, we can observe the existence of contradictory norms in different isolated societies/countries. For example, the "rule of the road" regarding which side of road to drive is different in different countries. In this section, we study how this phenomenon can be replicated in our collective learning framework. We consider two groups of agents that have high intra-group interaction frequency but rare inter-group interaction frequency. We are interested in investigating whether different norms can be emerged between these two isolated subgroups and when the same norm can be evolved among them. For our study, we consider two subpopulation of equal size (500 agents per subpopulation) and use the CGHP game and collective learning EV-l for illustration purpose. Similar results can be observed across all the previous games we considered. For each agent, during its interaction with its neighbors, it interacts with each of its neighbors with probability 1-p, while interacting with one agent randomly selected from another subpopulation with probability p.

Figure 9 shows the dynamics of the frequency of converging to a consistent norm (a,a) or (c,c) when the interaction frequency between subpopulations is increased gradually under the four network structures. Initially when there is no interaction between the two subpopulations, these two subpopulations evolve their norms independently. Thus we can see that the total frequency of

converging to a consistent norm is 0.5. For 50% of the runs, the two subpopulations evolve different norms. With the interaction frequency p increased, we can see that the frequency of converging to a consistent norm is increased gradually. For all four networks, the two subpopulations can always succeed in evolving towards a consistent norm when their inter-subpopulation interaction frequency p exceeds certain threshold. Besides, by comparing the results among the four topologies, we observe that random network performs best: It only requires 10% of random interaction between subpopulations to reach a consistent norm with 100% frequency. We hypothesize that it is due to the fastest information exchange speed in random network, since it allows an agent to interact with any other agents with equal probability.

4 PREVIOUS WORK

Until now, much research effort has been devoted to investigating the emergence of norms in agent societies through local interaction in the bottom-up manner. Shoham and Tennenholtz (1997) first investigated the norm emergence problem in agent society based on a simple and natural strategy—the highest cummulative reward (HCR). Sen and Airiau (2007) later proposed a social learning framework to investigate the emergence of social norms in a population of agents using three existing representative multiagent learning approaches in anti-coordination game. However, this framework only allows random interaction among agents and does not take into consideration the fact that in practice agents' interaction might be constrained by the underlying topologies they are situated in.

Later, a great deal of work (Sen and Sen 2010; Villatoro et al. 2009) extended this social learning framework by taking into consideration complex network topologies (i.e., the small-world and scale-free network) to model the underlying interaction of the agent society and investigated the influence of different network topologies and different system parameters (e.g., population size, neighborhood size) on the overall learning performance of converging to social norms. Villatoro et al. (2011) employed two social instruments (namely rewiring and observation) to facilitate the emergence of norms in agent society through dissolving the metastable subnorms. Simulation results showed that the combination of observation and rewiring enables agents to coordinate on norms by eliminating the subnorms efficiently. However, all these studies are based on the simple interaction protocol that each agent is only allowed to interact with one randomly chosen neighbor each round. In real-life situations, individuals may interact with multiple neighbors simultaneously and also make collective decisions based on the multiple available choices, which is not modeled in the social learning framework. Besides, only coordination game and anti-coordination game are considered in previous work, which cannot handle more complicated games such as penalty game and climbing game.

Yu et al. (2013, 2014) proposed two strategies (collective learning-l and collective learning-g) to promote the emergence of norms where agents are allowed to make collective decisions within networked societies. The authors focused on the "rule of the road" example modeled as a two-player coordination game and showed that their collective learning framework can promote faster norm convergence compared with social learning framework. However, their framework is designed only for the special case of coordination game. Similarly, Yu et al. (2016) recently propose a novel learning model to study the consensus formation problem among a population of agents. Each agent is allowed to adjust its behavior based on its own opinion and the guided one, which is generated following evolutionary game theory. However, in practice, the norm emergence problem can be much more complex and challenging, which may require agents to coordinate on different actions and also require agents to have the ability of overcoming high mis-coordination cost and the noise of the environment. All the above challenges cannot be handled by their proposed strategies.

Recently, a number of hierarchical learning strategies (Yu et al. 2015; Yang et al. 2016) have also been proposed to improve the norm emergence rate for the huge action space problem. For example, in Yang et al. (2016), subordinate agents report their information to their supervisors, while supervisors can generate instructions (rules and suggestions) based on the information collected from their subordinates. Subordinate agents heuristically update their strategies based on both their own experience and the instructions from their supervisors. Experimental results show that introducing hierarchical organization into agent society can significantly increase the norm emergence rate.

Fixed strategy agents play a critical role in influencing the direction of norm emergence and have received wide range of attention in previous work (Sen and Airiau 2007; Marchant et al. 2015; Franks et al. 2013; Griffiths and Anand 2012; Genter et al. 2015; James et al. 2015). Fixed strategy agents are those who always select the same action regardless of its efficiency or others' choices. Previous work has shown that inserting relatively small numbers of fixed strategy agents can significantly influence much larger populations when placed in networked social learning framework.

There also exist another line of work (Matlock and Sen 2009; Griffiths and Luck 2010a, 2010b; Griffiths 2008; Hao and Leung 2013; Chan et al. 2015) investigating effective mechanism (e.g., tag-based mechanism) to facilitate norm emergence in selfish agent-based systems. Griffiths and Luck (2010b) investigated the question of how a suitable set of norms can be established in a group of selfish agents. They proposed a tag-based interaction protocol and explored different factors that affect the adoption and longevity of cooperative norms in tag-based interaction environment. McDonald and Sen (2009, 2007) proposed three new tag mechanisms facilitate cooperation among selfish agents. The first tag mechanism is called Tag matching patterns (one and two-sided). Each agent is equipped with both a tag and a tag-matching string, which determines the interaction pattern among agents. The second mechanism is payoff sharing mechanism, which requires each agent to share part of its payoff with its opponent. This mechanism is shown to be effective in promoting socially optimal outcomes in both the prisoner dilemma game and the anti-coordination game; however, it can be applied only when side-payment is allowed. The last mechanism they propose is called a paired reproduction mechanism. It is a special reproduction mechanism that makes copies of matching pairs of individuals with mutation at corresponding place on the tag of one and the tag-matching string of the other at the same time. The purpose of this mechanism is to preserve the matching between this pair of agents after mutation to promote the survival rate of cooperators. Simulation results show that this mechanism can help sustaining the percentage of agents coordinating on socially optimal outcomes at a high level in both the prisoner dilemma and anti-coordination games. To summarize, all the work in this line focuses on the non-cooperative environment and investigate effective incentive mechanism to induce selfish agents towards socially optimal norms. However, our line of work focuses on the cooperative environment and how to overcome the stochasticity and dynamics in distributed environment to achieve efficient norm emergence.

5 CONCLUSION AND FUTURE WORK

We proposed two novel learning strategies for agents to converge to consistent norms through local interaction in different distributed multiagent environments under the collective learning framework. Extensive simulation shows that collective learning EV-l and EV-g can enable agents to reach consistent norms more efficiently and in a wider variety of games compared with previous approaches under both collective learning and social learning framework. The influence of different system parameters is also investigated. We find that the topology itself has no

significant influence on the norm emergence and norm emergence rate is increased with the increase of population size, action space, and the decrease of neighborhood size.

We empirically showed that the heuristic learning strategies under the collective learning framework is robust to the underlying network topology in terms of norm emergence, it would be interesting to further investigate how this desirable property can be theoretically verified. Another worthwhile direction is to apply the heuristic learning strategies to other multiagent coordination problems such as coordination in cooperative games under the collective learning framework. Last, it would also be worthwhile to investigate how the norm emergence rate can be further improved using abstraction techniques such as hierarchical multiagent learning (Makar et al. 2001).

ACKNOWLEDGEMENT

This work is partially supported by Tianjin Research Program of Application Foun- dation and Advanced Technology (No.: 16JCQNJC00100), National Natural Science Foundation of China (No.: 71502125, 61672358, 61502072), Fundamental Research Funds for the Central Universities of China (No.: DUT16RC(4)17). Additional corre- sponding author: Zhong Ming (mingz@szu.edu.cn).

REFERENCES

T. Agotnes and M. Wooldridge. 2010. Optimal social laws. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, Volume 1*. 667–674.

A. Alexander, K. Lloyd, J. Pitt, and M. Sergot. 2005. A protocol for resource sharing in norm-governed ad hoc networks. In *Declarative Agent Languages and Technologies II*. 221–238.

A. Alexander, S. Marek, and P. Jeremy. 2009. Specifying norm-governed computational societies. *ACM Trans. Comput. Logic* 10, 1 (2009), 1.

A. L. Barabási, R. Albert, and H. Jeong. 2000. Scale-free characteristics of random networks: The topology of the world-wide web. *Physica A* 281, 1 (2000), 69–77.

C. K. Chan, J. Y. Hao, and H. F. Leung. 2015. Reciprocal social strategy in social repeated games. In *Proceedings of International Conference on Tools with Artificial Intelligence*. 966–973.

N. Criado, E. Argente, A. Garrido, J. A. Gimeno, F. Igual, V. Botti, P. Noriega, and A. Giret. 2011. Norm enforceability in electronic institutions? In *Coordination, Organizations, Institutions, and Norms in Agent Systems VI*. 250–267.

H. Franks, N. Griffiths, and A. Jhumka. 2013. Manipulating convention emergence using influencer agents. *Auton. Agents Multi-Agent Syst.* 26, 3 (2013), 315–353.

K. Genter, S. Zhang, and P. Stone. 2015. Determining placements of influencing agents in a flock. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. 247–255.

N. Griffiths. 2008. Tags and image scoring for robust cooperation. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems, Volume 2*. 575–582.

N. Griffiths and S. S. Anand. 2012. The impact of social placement of non-learning agents on convention emergence. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*. 1367–1368.

N. Griffiths and M. Luck. 2010a. Changing neighbours: Improving tag-based cooperation. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*. 249–256.

N. Griffiths and M. Luck. 2010b. Norm emergence in tag-based cooperation. In *Proceedings of the 8th European Workshop on Multi-Agent Systems*.

J. Y. Hao and H. F. Leung. 2013. Achieving social optimality with influencer agents. In *Complex Sciences*. Springer, 140–151.

M. James, N. Griffiths, and M. Leeke. 2015. Convention emergence and influence in dynamic topologies. In *Proceedings of International Conference on Autonomous Agents and Multiagent Systems*. 1785–1786.

S. Kapetanakis and D. Kudenko. 2002. Reinforcement learning of coordination in cooperative multiagent systems. In *Proceedings of the 18th National Conference on Artificial Intelligence*. 326–331.

R. Makar, S. Mahadevan, and M. Ghavamzadeh. 2001. Hierarchical multi-agent reinforcement learning. In *Proceedings of the 5th International Conference on Autonomous Agents*. 246–253.

J. Marchant, N. Griffiths, and M. Leeke. 2015. Manipulating conventions in a particle-based topology. In *Proceedings of the Coordination, Organizations, Institutions and Norms in Agent Systems Workshop*.

M. Matlock and S. Sen. 2007. Effective tag mechanisms for evolving coordination. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*. ACM, 251.

M. Matlock and S. Sen. 2009. Effective tag mechanisms for evolving cooperation. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems*. 489–496.

M. Mihaylov, K. Tuyls, and A. Nowé. 2014. A decentralized approach for convention emergence in multi-agent systems. *Autonomous Agents and Multi-Agent Systems* 28, 5 (2014), 749–778.

J. Morales, M. Lopez-Sanchez, J. A. Rodriguez-Aguilar, M. Wooldridge, and W. Vasconcelos. 2013. Automated synthesis of normative systems. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems*. 483–490.

R. Polikar. 2006. Ensemble based systems in decision making. *Circ. Syst. Mag.* 6, 3 (2006), 21–45.

S. Redner. 1998. How popular is your paper an empirical study of the citation distribution. *Eur. Phys. J. B* 4, 2 (1998), 132–134.

A. Réka and A. L. Barabási. 2002. Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 1 (2002), 47–97.

O. Sen and S. Sen. 2010. Effects of social network topology and options on norm emergence. In *Coordination, Organizations, Institutions and Norms in Agent Systems V*. 211–222.

S. Sen and S. Airiau. 2007. Emergence of norms through social learning. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence*. 1507–1512.

Y. Shoham and M. Tennenholtz. 1997. On the emergence of social conventions: Modeling, analysis and simulations. *Artif. Intell.* 94(1-2) (1997), 139–166.

D. Villatoro, J. Sabater-Mir, and S. Sen. 2011. Social instruments for robust convention emergence. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*. 420–425.

D. Villatoro, S. Sen, and J. Sabater-Mir. 2009. Topology and memory effect on convention emergence. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, Volume 02*. 233–240.

X. F. Wang and G. R. Chen. 2003. Complex networks: Small-world, scale-free and beyond. *IEEE Circ. Syst. Mag.* 3, 1 (2003), 6–20.

T. P. Yang, Z. P. Meng, J. Y. Hao, S. Sen, and C. Yu. 2016. Accelerating norm emergence through hierarchical heuristic learning. In *Proceedings of the 22th European Conference on Artificial Intelligence (ECAI'16)*. 1344–1352.

H. P. Young. 1996. The economics of convention. *J. Econ. Perspect.* 10, 2 (1996), 105–122.

C. Yu, H. T. Lv, F. H. Ren, H. L. Bao, and J. Y. Hao. 2015. Hierarchical learning for emergence of social norms in networked multiagent systems. In *AI 2015: Advances in Artificial Intelligence*. Springer, Berlin, 630–643.

C. Yu, G. Z. Tan, H. T. Lv, Z. Wang, J. Meng, J. Y. Hao, and F. H. Ren. 2016. Modelling adaptive learning behaviours for consensus formation in human societies. *Sci. Rep.* 6 (2016).

C. Yu, M. J. Zhang, and F. H. Ren. 2014. Collective learning for the emergence of social norms in networked multiagent systems. *IEEE Trans. Cybernet.* 44, 12 (2014), 2342–2355.

C. Yu, M. J. Zhang, F. H. Ren, and X. D. Luo. 2013. Emergence of social norms through collective learning in networked agent societies. In *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems*. 475–482.

Received November 2015; revised January 2017; accepted July 2017